EINSTEIN'S RIDDLE AS A TOOL FOR PROFILING STUDENTS

Vildan Özeke[1] and Gökhan Akçapınar[2]
[1]*Gaziosmanpaşa University, Turkey, Faculty of Education, Department of Computer Education and Instructional Technology*
[2]*Hacettepe University, Turkey, Faculty of Education, Department of Computer Education and Instructional Technology*

ABSTRACT

There are many computer games, learning environments, online tutoring systems or computerized tools which keeps the track of the user while learning or engaging in the activities. This paper presents results from an exploratory study and aims to group students regarding their behavior data while solving the Einstein's riddle. 45 undergraduate students were given this logic puzzle as a complex cognitive task without any time limitation. After completing the task, they were asked to report their mental effort. While grouping the similar students, cluster analysis with X-Means algorithm was used. Features such as task performance, puzzle's difficulty levels, item movements inside and between the puzzle's sections, duration and a total number of incorrect moves were used while grouping students. At the end of the lab session, six out of forty-five participants solved the puzzle and find the correct answer, on the other hand, other students reached different completion levels. Based on cluster analysis students grouped into three different clusters, Cluster_0, Cluster_1 and Cluster_2. Cluster_2 was the successful group with the highest score, lowest moves and errors, medium level of mental effort in the shortest time period. Cluster_0 had the medium level of success with highest moves and errors, the highest level of mental effort in the highest time period. Cluster_1 was the least successful group with the lowest score, medium level of moves and errors and lowest level of mental effort.

KEYWORDS

Student profiling, clustering, educational data mining, complex cognitive task, mental effort, logic puzzle

1. INTRODUCTION

Computerized tools allow researchers to collect all kind of data about interaction between the user and the system in an unobtrusive way without disturbing the user (Khenissi et al., 2015). By analyzing data - obtained from these tools - with the help of data mining techniques can be used in different areas. One of this application area is grouping users based on their similarities (personal preferences, characteristics, etc.). Grouping similar users (user profiling) can be used for many purposes in many domains such as business for marketing strategies and personalized advertising, or gaming industry for categorizing the players (e.g. casual players, weekenders, social players, big spenders, decorators etc.) (Bienkowski et al., 2012). In educational setting, user profiling has been used to increase the learning performances and effectiveness when organizing adaptive/individualized learning environments (Bienkowski et al., 2012), in intelligent tutoring systems to present adaptive interaction support, to create online study groups according to clustering results (Kardan and Conati, 2011).

Learner profiling is usually conducted by analyzing the learner logs retrieved from e-learning environments (Wang and Liao, 2011, Akçapınar et al., 2016) or educational games (Hawlitschek and Köppen, 2014) with data mining and machine learning algorithms (Bienkowski et al., 2012). Apart from these studies, in the present study we used a logic puzzle (Einstein's Riddle) as a data collection tool with the aim of grouping similar students. Correlation between students' self-reported mental effort scores and log based features was also investigated for each cluster.

ISBN: 978-989-8533-55-5 © 2016

1.1 Einstein's Riddle as a Complex Cognitive Task

Solving a logic puzzle is required to use high-level cognitive skills (e.g. reasoning, problem-solving, analytical/critical thinking, etc.) and cognitive processes (e.g. attention, coding, storing, mental shifting, mental effort etc.) together. Cognitive competence of a person was formed by cognitive skills which allow individuals to distinguish objects, events or stimuli, to identify and categorize the concepts, to build issues, rules and make them "problem solvers" with high-level mental processing (Otero et al., 2012).

Therefore, these puzzles can be used to assess cognitive skills of students and to profile their cognitive traits and situations. Moreover, we can also categorize them as "*complex cognitive task*" according to Wood's (1986) task complexity definition. As stated by Wood, there are three types of task complexity: a) component, b) coordinative, and c) dynamic complexity. The component complexity is a function of the number of information cues to be processed and the number of acts which need to be executed during the task performance (e.g. chess game). As the numbers increase the task complexity increases. The coordinative complexity related to the power of relationships (strong/weak) between task inputs (information cue & required acts) and task products. While the learner performed acts in one part of the task, several other acts need to be performed concurrently (e.g. radio assembly). The dynamic complexity refers to the change in time of both task inputs/outputs, and the relationships between them (e.g. decision making). For example, air-traffic controlling used to be known as complex cognitive tasks which include these three types of task complexity. According to Wood (1986), total complexity of a task derives from these three types of complexity.

There are rules, hints and puzzle items as information cues in Einstein's riddle (see Figure 1), and learners have to act between areas by dragging and dropping, clicking and checking the boxes simultaneously. All these actions can be considered as component and coordinative complexity. Some hints may need to be returned and to read again for solving the puzzle and the participant need to uncheck/recheck it for the upper-level boxes and this situation can be considered as dynamic complexity.

Finally, it is important to track and record the cognitive processing of a learner while s/he was engaging with a cognitive task. The aim of this study is to determine user behavior and to name their profiles by using a computerized complex cognitive task.

2. METHOD

2.1 Study Design

This is an exploratory design study conducted with 45 undergraduate students (24 females and 21 males) in the Computer Education and Instructional Technology (CEIT) Department in a state-funded university in Turkey. Participants' ages ranged between 20 and 27 with the mean of 21.92 (SD = 1.47). Each student has completed the computer-based task on their own without any time limitation. The minimum and maximum duration for all participants were between 5 and 39 minutes (M = 19.88, SD = 8.02). All participants were a volunteer and dealt with the task in a computer laboratory. The instructions were given by the authors and there was no time limitation for completing the task. The participants were informed about they have a right to quit the task anytime they want. After completing the task, they have asked to self-reported their amount of mental effort.

2.2 Material

The complex cognitive task used in a study is known as Einstein's five-house riddle (see Appendix). There is no evidence about who invented the puzzle but it is very popular among logic puzzles. In general, different kind of animals and cigarette brands are used as a puzzle items. Because of educational concern of the study, we have changed cigarette brands with car brands. The authors developed a computerized version of this puzzle. The tool was developed using C# programming language on Windows Presentation Foundation (WPF) platform. Event based logging system was also implemented. Therefore, the tool is able to log all events (check, uncheck, move, etc.) during the session with a timestamp. Following information was given to

the students as part of the puzzle: rules to be considered, fifteen hints, and a question. They can click the checkbox in front of the hint when they think they have used that information to solve the puzzle and it turns red as seen in the "hints" area in Figure 1. They can drag and drop each of the "puzzle items" over the "solution matrix" (from Section A to Section B), they can also move items inside the section A and section B as they prefer. In the "hints" section there are two more buttons: "Restart" (works as a reset) and "Finish" (to use after completing the task or when they would like to quit).

Figure 1. A screenshot from the computerized version of Einstein's five-house riddle

Merrill (2006a, 2006b) suggests researchers to create rubrics for the evaluation of task performance, and to calculate different performance levels at the end of the process. According to Merrill (2006a, 2006b) the number of transactions/steps where the complexity ascends should be stated in order to be able to assess the performance levels for complex tasks. A rubric scale was developed to calculate the performance score of participants (over 100 points). There are four levels to get the final answer. Level 1 has the basic boxes to fill. While the level increases the puzzle requires more attention and the scores of the boxes increase. There are 25 boxes to fill while solving the puzzle and all truly filled box adds some points to the participant's overall score. While calculating the total score of the participants the scale given in Table 1 was used.

Table 1. Calculating the performance score over 100 points

Difficulty level	Count of boxes	Points for each	Total Point
Level 1	3 boxes	1 point	3
Level 2	8 boxes	3 points	24
Level 3	6 boxes	5 points	30
Level 4	7 boxes	6 points	42
The answer	1 box	1 point	1

2.3 Rating Scale Mental Effort (RSME)

Zijlstra (1993) gave some visual search tasks to his driver participants and then measured their mental effort in his dissertation. The vertical scale has a range between 0-150 points from "hardly any effort" to "extreme effort". According to Zijlstra (1993; pp.34-35) the amount of the effort depends on the following three variables as a) the demands of the task, b) subject's available performance potential and c) the duration of the activity (time-on-task). The reliability of the scale was r = .81 in a laboratory setting and r = .71 in a real work setting. The participants self-reported their mental effort after finishing the task.

2.4 Features

Computerized version of the puzzle is able to log every action done by the students. Each session was logged in separate log files. Student's ID was used as a unique identifier to join different data sources. Analysis data generated automatically by the developed preprocess tool. The dataset used in the cluster analysis consisted of 45 students' usage data with 11 features. Four of them related to student's moves inside the game. Four of them related to student's achievements across the different levels. One is a number of error (incorrect placements) done by the student. One is game duration. And the last one is showed the highest score achieved by the student during the game. List of the features and their explanations can be seen in Table 2.

Table 2. Description of features

Feature	Description
AA[a]	Total number of moves inside section A (item area)
AB[a]	Total number of moves from section A to B
BA[a]	Total number of moves from section B to A
BB[a]	Total number of moves inside section B (solution area)
L1	Total number of correct placements in Level 1 difficulty
L2	Total number of correct placements in Level 2 difficulty
L3	Total number of correct placements in Level 3 difficulty
L4	Total number of correct placements in Level 4 difficulty
Duration	Total time (minutes) spent in puzzle
Error	Total number of incorrect moves
Score	Highest score achieved during the session

[a] Section A and Section B can be seen in Figure 1.

2.5 Data Analysis

Data analysis was performed by using cluster analysis. Cluster analysis is widely used in educational data mining studies to identify similar groups of students. X-Means was used as a clustering algorithm; it is a modified version of the K-Means. Unlike K-Means algorithm, X-Means does not need to perform any clustering a-priori. It directly finds the optimum number of clusters from the data by using Bayesian Information Criteria (BIC) (Pelleg and Moore, 2000). X-Means algorithm was selected since we have no a priori knowledge about the number of hidden groups in the data. Cluster analysis was performed RapidMiner data mining software with process given in Figure 2. Cosine similarity measure was used as a distance metric and all features were converted to z-scores before the analysis.

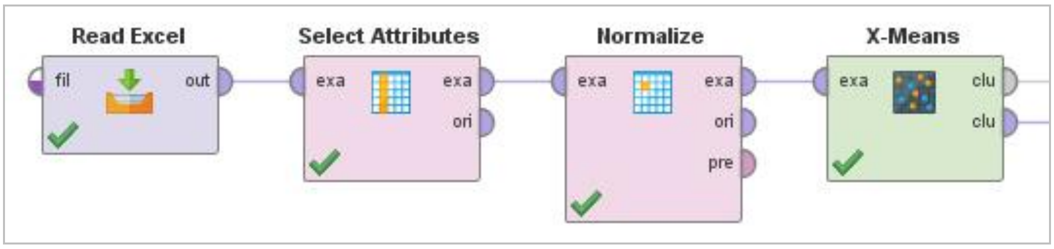

Figure 2. Cluster analysis process

3. RESULTS

As a result of the cluster analysis three different groups of students were obtained, Cluster_0, Cluster_1 and Cluster_2. According to cluster means given in Table 3 and Figure 3, Cluster_0 spent more time during the task than Cluster_1 and Cluster_2. Cluster_0 also had the most moves inside and between the sections (item and solution area), had the most errors but got average level of success. We can infer that Cluster_0 really did their best to accomplish the task. On the other hand, Cluster_2 was the one who spent the minimum time and had the minimum moves with minimum errors. Cluster_2 got the highest success in contrast with Cluster_1. However, Cluster_1 was the least successful group with medium errors. They had very close values to Cluster_2 in terms of moves between sections with only a distinct difference that Cluster_1 had more moves inside section A than Cluster_2. Their performance duration was similar.

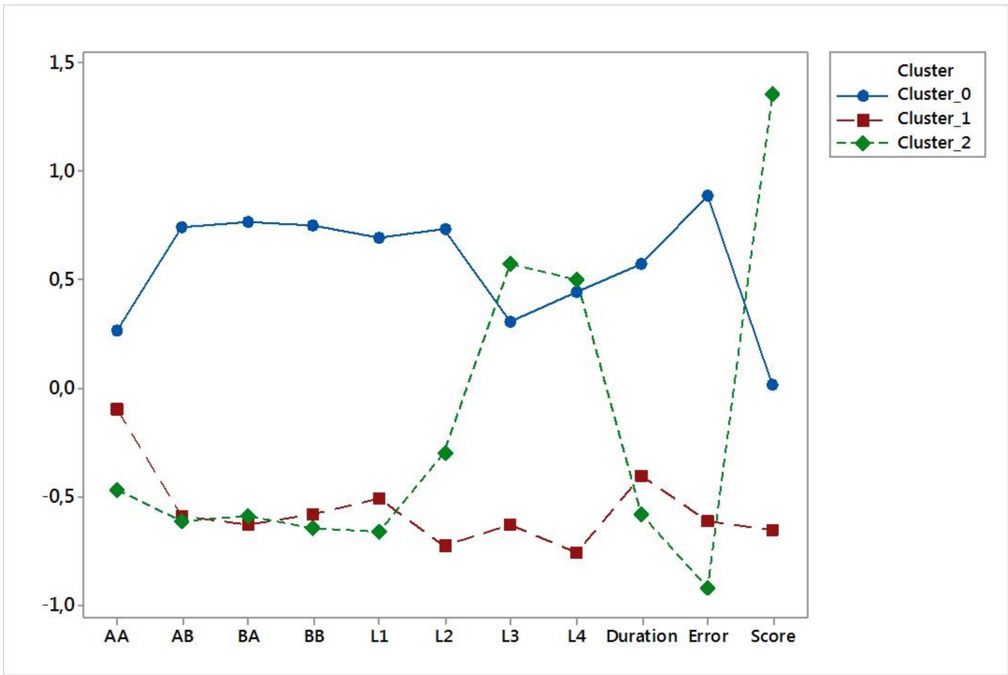

Figure 3. Normalized cluster centroids

Six of the participants placed all of the 25 items correctly and find the answer. Four of them are in the Cluster_2, two of them in Cluster_0. Other participants, however, reached the different level of completion. Their scores range from 8 to 93 (M = 43.0, SD = 21.0).

Table 3. Cluster means and standard deviations for all features

Feature	Cluster_0 (n = 18)	Cluster_1 (n = 18)	Cluster_2 (n = 9)
AA	29.4 (38.5)	15.4 (24.1)	4.8 (6.9)
AB	59.9 (18.8)	34.1 (9.0)	34.3 (10.5)
BA	25 (13.1)	5.8 (6.3)	5.4 (6.2)
BB	61.5 (32.3)	21.9 (12.2)	21.3 (15.8)
L1	7.6 (2.6)	4.4 (1.8)	4.2 (2.2)
L2	14.6 (6.7)	5.3 (2.5)	8.3 (2.3)
L3	6.2 (4.6)	2.2 (2.0)	6.7 (1.2)
L4	6.6 (3.7)	2.1 (1.5)	6.6 (1.4)
Duration	25.3 (5.9)	16.7 (7.7)	15.4 (6.4)
Error	111.4 (30.2)	47.8 (22.5)	35.3 (21.3)
Score	50.7 (22.9)	31.7 (15.5)	88.1 (14.1)

In terms of mental effort, students in Cluster_0 self-reported more average mental effort (M = 77.22, SD = 30.64) than those in Cluster_2 (M = 70.55, SD = 22.70) and Cluster_1 (M = 56.94, SD= 21.77). Cluster_1 reported the minimum mental effort. A Pearson product-moment correlation coefficient was computed for each cluster to assess the relationship between the log based features and students' self-reported RSME scores. When results in Table 4. were examined, statistically significant correlations can be observed only in Cluster_0. In this group, there was a moderate positive correlation between RSME and Total Moves variables. For this group increases in moves were correlated with perceived mental effort. There was also a moderate positive correlation between RSME and Score variables.

Table 4. Correlations between RSME scores and log variables for each cluster

	Rating Scale Mental Effort (RSME) Scores		
Feature	Cluster_0 (n = 18)	Cluster_1 (n = 18)	Cluster_2 (n = 9)
Total Moves	.480*	-.195	.376
Total Errors	-.077	.051	.258
Score	.499*	-.078	.364
Duration	.443	.123	.188

*Correlation is significant at the 0.05 level (2-tailed).

4. CONCLUSION

In this paper, we aimed to group similar students regarding their behavior data while solving the complex cognitive task (Einstein's riddle). The cluster analysis of the behavioral data revealed three different groups, Cluster 0, 1 and 2. We found that students who took the shortest time and made less moves in solving the puzzle obtained the highest scores (Cluster_2) while students who took the longest time and more moves obtained moderate scores (Cluster_0). Obviously one of the most interesting findings obtained here is that, although Cluster_1 and Cluster_2 are most distinct clusters in terms of performance, students in these clusters has a lot in common. For instance, they both have the minimum number of interactions; both spend a shorter time on the puzzle when compared to Cluster_0. If we didn't include students' performance in cluster analysis, most of the students in these clusters could be assigned to the same cluster. This finding shows the importance of the performance metrics in educational data mining studies while extracting student profiles.

Paas and Merriënboer (1993) formulized performance and mental effort and named as "mental efficiency". If an individual gets higher performance with the lowest mental effort they said "higher mental efficiency", despite that lowest performance with the highest mental effort was called as "lowest mental efficiency" (Paas et al., 2003). Sometimes they used speed (task duration while completing the task) as a secondary metric in addition to the mental effort. The complexity and mental effort are generally correlated with each other, mental effort usually be treated as indices of cognitive load (Clark and Elen, 2006). In terms of mental effort scale scores, students in Cluster_0 reported highest mental effort when compared to students in other clusters. One possible explanation of this could be as students in this cluster take the task seriously and tried to do their best. It is a limitation of this study not to use other tests to measure attention/sustained attention. In further studies such cognitive preferences should be carried out to the research design.

For an optimal solution 25 moves from Section A (item area) to B (solution matrix) could be enough to solve the puzzle however average moves of the students in Cluster_0 seven times higher than that and most of these moves occurred in the solution area while changing the item one place to another. Students who have higher performances have less mouse clicks and moves than others. Unintentional mouse movements may be used for measuring the degree of concentration or frustration of learner (Khenissi et al., 2015).

According to Alloway et al. (2009), learners with low-level working memory give up the complex tasks without struggling with it. In our study, the lowest performance of Cluster_1 may be a result of participants' low level working memory, since, Cluster_1 has similar moves and duration like Cluster_2, yet, in terms of performance there is a huge difference between them. In further studies obtained clusters can be analyzed in terms of working memory capacity.

This study showed to possible usage of the logical reasoning puzzle as a student profiling tool. It may be the first time to use logical reasoning puzzles to measure the cognitive skills of learners. In the literature there

were examples which used games and learning environments/materials for measuring. The path to solving the puzzle and the decisions the students make may be related to learners' cognitive skills (Taiyu and Kinshuk, 2009). However, further studies are needed to understand the characteristic of these students in more details in terms of more cognitive preferences. These puzzles also can be used while determining the at-risk students (as Cluster_1 in our study). Teachers provide individualized advice to the learners according to their clusters (Alfredo et al., 2010). We can use these puzzles to create learner profiles in adaptive systems. Furthermore, as mentioned by Rodrigo et al. (2008) obtained results can in the future be used to design a prediction model that is capable of detecting different learner profiles. While predicting the learner behavior and describing their peculiarities we can create models by using the captured log files and recorded data structures as the trails of learner actions (Jovanovic et al., 2012).

REFERENCES

Akçapinar, G., Altun, A. & Aşkar, P. 2016. Çevrimiçi Öğrenme Ortamındaki Farklı Öğrenci Profillerinin Kümeleme Yöntemi ile Belirlenmesi. *Eğitim Teknolojisi Kuram ve Uygulama,* 6**,** 46-64.

Alfredo, V., Félix, C. & Àngela, N. 2010. Clustering Educational Data. *Handbook of Educational Data Mining.* CRC Press.

Alloway, T. P., Gathercole, S. E., Kirkwood, H. & Elliott, J. 2009. The cognitive and behavioral characteristics of children with low working memory. *Child Dev,* 80**,** 606-21.

Bienkowski, M., Feng, M. & Means, B. 2012. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Washington, D.C.

Clark, R. E. & Elen, J. 2006. When less is more: Research and theory insights about instruction for complex learning. *In:* Elen, J. & Clark, R. E. (eds.) *Handling Complexity in Learning Environments: Theory and Research.* Oxford, UK: Elsevier.

Hawlitschek, A. & Köppen, V. Analyzing Player Behavior in Digital Game-Based Learning: Advantages and Challenges. 8th European Conference on Games Based Learning: ECGBL2014, 2014 Berlin, Germany. Academic Conferences and Publishing International, 199-206.

Jovanovic, M., Vukicevic, M., Milovanovic, M. & Minovic, M. 2012. Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study. *International Journal of Computational Intelligence Systems,* 5**,** 597-610.

Kardan, S. & Conati, C. A Framework for Capturing Distinguishing User Interaction Behaviours in Novel Interfaces. The 4th International Conference on Educational Data Mining (EDM 2011), 2011 2011. Eindhoven, the Netherlands.

Khenissi, M. A., Essalmi, F., Jemni, M. & Kinshuk 2015. Learner Modeling Using Educational Games: A Review of the Literature. *Smart Learning Environments,* 2**,** 1-16.

Merrill, M. D. 2006a. Hypothesized performance on complex tasks as a function of scaled instructional strategies. *In:* Elen, J. & Clark, R. E. (eds.) *Handling Complexity in Learning Environments: Theory and Research.* Oxford, UK: Elsevier.

Merrill, M. D. 2006b. Levels of instructional strategy. *Educational Technology,* 46**,** 5-10.

Otero, W. I., Petch, J. & Catapan, A. H. 2012. Developing high-level cognitive skills in distance education. *InterSciencePlace: Revista Cientifica Internacional,* 20**,** 41-57.

Paas, F., Tuovinen, J. E., Tabbers, H. & Van Gerven, P. W. M. 2003. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist,* 38**,** 63-71.

Paas, F. G. W. C. & Merriënboer, J. J. G. V. 1993. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 35**,** 737-743.

Pelleg, D. & Moore, A. W. 2000. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *Proceedings of the Seventeenth International Conference on Machine Learning.* Morgan Kaufmann Publishers Inc.

Rodrigo, M., Anglo, E., Sugay, J. & Baker, R. Use of unsupervised clustering to characterize learner behaviors and affective states while using an intelligent tutoring system. International Conference on Computers in Education, 2008. 57-64.

Taiyu, L. & Kinshuk 2009. Cognitive Profiling in Life-Long Learning. *In:* Patricia, L. R., Gary, A. B., Judith, V. B., Caroline, H., Lorraine, J. & Karen, D. S. (eds.) *Encyclopedia of Distance Learning, Second Edition.* Hershey, PA, USA: IGI Global.

Wang, Y.-H. & Liao, H.-C. 2011. Data mining for adaptive learning in a TESL-based e-learning system. *Expert Systems with Applications,* 38**,** 6480-6485.

ISBN: 978-989-8533-55-5

Wood, R. E. 1986. Task complexity: Definition of the construct. *Organizational behavior and human decision processes*, 37**,** 60-82.

Zijlstra. 1993. *Efficiency in work behaviour: A design approach for modern tools (Unpublished doctorate dissertation).* TU Delft, Delft University of Technology.

APPENDIX

Einstein's riddle

The situation

1. There are 5 houses in five different colors.
2. In each house lives a person with a different nationality.
3. These five owners drink a certain type of beverage, drive a certain brand of car and keep a certain pet.
4. No owners have the same pet, drive the same brand of car or drink the same beverage.

The question is: *Who owns the fish?*

Hints

1. The Brit lives in the red house
2. The Swede keeps dogs as pets
3. The Dane drinks tea
4. The green house's owner drinks coffee
5. The green house is on the left of the white house
6. The person who drives Porsche rears birds
7. The owner of the yellow house drives Ferrari
8. The man living in the center house drinks milk
9. The Norwegian lives in the first house
10. The man who drives BMW lives next to the one who keeps cats
11. The man who keeps horses lives next to the man who drives Ferrari
12. The owner who drives Audi drinks mineral-water
13. The German drives Alfa-Romeo
14. The Norwegian lives next to the blue house
15. The man who drives BMW has a neighbor who drinks water